Niveau(x) d'enseignement: 7 Numéro EDUQ : 0470093
Centre d'intérêt: 570

Brunet, Richard.

L'analyse psychométrique du test PERPE / Richard Brunet. -- [Montréal] : PERPE, [1971?].
12 f. ; 26 cm.

Document: 0671-02.

Descripteur(s): *Evaluation de l'enseignant *Evaluation par l'étudiant *Satisfaction de l'étudiant *Attitude de l'enseignant *Relations maître-élèves *Evaluation de tests *Test *Psychométrie *Collégial (Niveau)

Identificateur(s) de nom: *PERPE (Test)

Après une description de la section "recherches psychométriques" du projet PERPE (Perceptions étudiantes de la relation professeur-étudiants), ce texte présente l'aspect psychométrique du test PERPE (son contenu, sa structure et ses qualités psychométriques), insistant à la fois sur les travaux qui ont mené à l'instrument actuel et sur les projets visant à assurer et à maintenir les hautes qualités scientifiques de ce test. Pour le contenu, on examine l'origine de 61 dimensions du test, la définition des points de l'échelle de chacune des dimensions et le regroupement des dimensions en sections. Pour la structure de l instrument, il s'agit de l'étude du construit insatisfaction, des problèmes de normalisation, de la purification des données et de la hiérarchisation des dimensions. Enfin, les qualités psychométriques examinées sont la fidélité et la validité du test.

Type de document: r

EDUQ 8470093

42

L'ANALYSE PSYCHOMÉTRIQUE DU TEST PERPE

Richard Brunet

Service PERPE
Document 0671-02

Le but de cet exposé étant de présenter l'aspect psychométrique du test PERFE, nous examinerons dans les minutes qui vont suivre, le contenu, la structure et les qualités psychométriques de l'instrument, insistant à la fois sur les travaux qui ont mené à l'instrument actuel et sur les projets visant à assurer et à maintenir les hautes qualités scientifiques de ce test.

Avant d'aller plus loin, il importe de préciser les principaux objectifs de la section 'Recherches psychométriques'. Ceux-ci sont au nombre de quatre: le premier consiste à confirmer les qualités et à identifier les faiblesses du test PERFE comme instrument scientifique de mesure de la situation d'enseignement au collégial, le second à améliorer l'instrument, surtout en ce qui a trait au Dossier de résultats transmis au professeur. Le troisième objectif consiste à maintenir constantes les hautes qualités scientifiques de l'instrument en l'adaptant régulièrement à mesure qu'évoluera la pédagogie à l'intérieur du réseau collégial. Ce même objectif sera visé lorsqu'il s'agira d'étendre l'instrument à une population différente. Enfin, le quatrième objectif vise à étendre le champ des réalités mesurées à l'intérieur de la situation d'enseignement par un programme de construction d'instruments nouveaux. Ainsi, quelle que soit la matière enseignée ou son approche pédagogique, un professeur trouverait éventuellement parmi nos instruments, celui ou ceux qui lui donneraient le portrait le plus fidèle et le plus complet de son enseignement.

OBJECTIFS DE LA SECTION 'RECHERCHES PSYCHOMÉTRIQUES'

1. Confirmer les qualités et identifier les faiblesses du test.
2. Améliorer l'instrument.
3. Adapter le test à l'évolution de la situation d'enseignement.
4. Etendre le champ des réalités mesurées.

Ces objectifs, si modestes qu'ils soient, sont rendus possibles par la nouvelle situation qui prévaut au Service PERT; en effet, d'une part, le nombre de chercheurs attachés à la Section 'Recherches psychométriques' vient de tripler, passant de un à trois,et d'autre part, depuis l'automne 1970, l'exploitation du test à la grandeur du réseau collégial nous fournit une somme considérable de données à analyser.

Comme vous pourrez le constater par la suite, les analyses psychométriques réalisées, soit partiellement, soit entièrement, apparaissent comme une portion assez mince par rapport à un ensemble de questions qui demeurent encore sans réponse. De plus, l'absence d'un plan directeur rendait par ailleurs difficile le choix des projets à réaliser prioritairement, d'autant que la hiérarchisation qu'on aurait pu faire de ces projets n'aurait pû être qu'arbitraire.

C'est dans ce contexte qu'est apparue la nécessité d'élaborer pour le secteur de la recherche psychométrique un programme intégré d'analyse définissant une séquence de projets à réaliser. Ce programme devrait nous permettre de réévaluer rétrospectivement les analyses déjà faites et d'organiser le travail des chercheurs pour les deux prochaines années.

Trois préoccupations majeures nous ont amené à accorder la priorité à un tel programme pour les années à venir:

a) au plan <u>scientifique</u>, une volonté de quitter le stade artisanal dans lequel nous étions pour accéder à une plus grande rigueur dans nos analyses;

b) au plan <u>administratif</u>, le désir de réviser complètement le Dossier personnel du professeur;

c) au plan <u>prospectif</u>, un projet de construction de nouveaux instruments.

En vue de répondre à ces objectifs, le programme d'analyse permettra l'identification des lacunes de notre instrument actuel et conduira à un raffinement de nos techniques de mesure. Ce programme consistera à prendre sys-

téoriquement chacune des opérations qui ont conduit à la construction du test PERPE, tel qu'il existe actuellement, et d'en vérifier la légitimité à l'aide de techniques rigoureuses. Les étapes du programme s'enchaînent donc suivant la logique même de la construction de notre instrument. C'est, en somme, ce dont je vais maintenant vous entretenir.

I - Contenu de l'instrument

Les travaux reliés au contenu de l'instrument se subdivisent en trois projets distincts, à savoir

- l'origine des 61 dimensions du test
- la définition des points de l'échelle de chacune des dimensions
- le regroupement des dimensions en sections.

Nous allons examiner successivement chacune de ces étapes.

1ère étape: Origine des 61 dimensions du test

Le test PERPE tire son origine d'un cours de psychologie sociale où les étudiants avaient pour tâche de clarifier les critères permettant de juger un cours. Puis, ceux-ci observèrent des professeurs et notèrent des faits de comportement qu'ils jugeaient importants. Plus de 600 fiches descriptives furent ainsi obtenues, fiches qui servirent de point de départ à la préparation de la première forme expérimentale du test.

Ce procédé empirique consistant à construire le test à partir de dimensions fournies par les étudiants a l'immense avantage de donner un questionnaire qui se situe dans le cadre de référence des personnes qui auront à y répondre, c'est-à-dire les étudiants. Ainsi, on s'assure que les dimensions à évaluer ont une réelle signification pour l'évaluateur.

Avant d'en arriver à la forme actuelle de l'instrument, où chaque dimension a acquis une valeur scientifique, le test PERFECT a subi cinq expérimentations contrôlées; au total, 97 dimensions ont été expérimentées et les 61 meilleures furent conservées sur la base de critères bien définis. Ces critères sont les suivants.

CRITÈRES D'UNE BONNE DIMENSION

1. La dimension doit refléter de l'insatisfaction puisque l'instrument mesure l'insatisfaction des étudiants.

2. La contradiction sur le sens de l'insatisfaction ne doit pas être trop élevée; en effet, un pourcentage de contradiction élevé indique que, parmi les étudiants insatisfaits, plusieurs jugent qu'il y a trop de quelque chose alors que les autres jugent qu'il n'y en a pas assez.

3. La dispersion des classes quant à l'insatisfaction moyenne manifestée doit être assez grande de façon à permettre une bonne discrimination entre les professeurs.

4. Les étudiants doivent s'entendre, à l'intérieur d'une même classe, sur l'évaluation de la réalité.

5. D'une classe à l'autre, l'évaluation de la réalité doit varier de façon à permettre une discrimination entre les professeurs.

6. Les étudiants doivent s'entendre sur leurs attentes face aux professeurs, celles-ci doivent être homogènes.

7. La dimension doit être pertinente; en d'autres mots, la proportion d'étudiants s'abstenant d'évaluer une dimension doit être faible.

8. La dimension doit évaluer l'insatisfaction des étudiants, mais sous un rapport autre que celui fourni par les autres dimensions.

Cette première étape, qui est maintenant presque terminée, consistait essentiellement à colliger une série de documents déjà existants; elle fera

bientôt l'objet d'une publication. Cependant, ces études devront être reprises périodiquement dans le but de s'assurer que les paramètres de chaque dimension seront stables d'une exploitation du test à l'autre.

2ième étape: La définition des points de l'échelle de chacune des dimensions

Le test PERPE utilise depuis son début des échelles en cinq points qui doivent répondre à certains critères statistiques. Premièrement, la moyenne des réponses ayant trait à l'évaluation de la réalité doit se trouver au point central de l'échelle, c'est-à-dire au point 3. Deuxièmement, l'évaluation de la réalité doit se distribuer normalement. Ainsi, on s'assure que l'échelle permet une évaluation des extrêmes de la réalité rendant plus facile la discrimination entre les professeurs. Les échelles ne se conformant pas à ces deux critères ont donc subi, à la suite des diverses expérimentations, des modifications visant à corriger ces défauts, ces modifications consistaient, la plupart du temps, à déplacer les bornes supérieures et inférieures de l'échelle.

Ces critères statistiques d'une bonne échelle ainsi que les analyses faites au cours des diverses expérimentations de l'instrument seront consignés dans un document qui sera publié. De plus, comme pour la première étape, on devra vérifier périodiquement les paramètres des échelles.

A cette partie historique viendront cependant s'ajouter de nouvelles analyses. Jusqu'à ce jour, en effet, nous avons mis l'accent sur la distribution des réponses des étudiants concernant la réalité perçue (réponses A) pour déterminer les points de l'échelle d'une dimension. Comme nous l'avons vu précédemment, le critère retenu était celui de la distribution normale des résultats. Or, il n'est pas certain que ce critère soit réellement valable pour toutes les dimensions; tout dépend en effet de la courbe observée des réponses B (attentes) sur la même dimension.

Aussi avons-nous défini un projet visant une classification des 61 dimensions du test selon la forme des distributions des réponses A et B.

nous devrions pouvoir définir ainsi des critères différentiels pour déterminer les points de l'échelle des diverses dimensions.

2ème étape: Le regroupement des dimensions en sections

Les 61 dimensions du test PERPE ont été regroupées en sections de façon à pouvoir faciliter l'interprétation des résultats. Historiquement, trois regroupements ont été tentés; le premier consistait à rassembler les dimensions selon leur contenu. Ces premières sections sont:

PREMIER REGROUPEMENT EN SECTION
1. La personnalité du professeur
2. Le contenu du cours
3. La présentation du cours
4. L'interaction professeur-étudiant.

Après avoir constaté que les sections ainsi formées regroupaient des dimensions parfois hétérogènes, on procéda à un nouveau regroupement en utilisant les catégories suivantes:

DEUXIÈME REGROUPEMENT	
Catégorie	Sous-catégorie
- A - Le message i.e. le cours	1. Ce qu'il est: caractéristiques liées au contenu du cours et aux techniques d'évaluation. 2. Comment il est transmis: organisation des contenus et des techniques pédagogiques.

- B - L'émetteur i.e. le professeur	1. C'est-à-dire: attitudes et traits de personnalité ayant des incidences pédagogiques. 2. Comment il transmet: caractéristiques immédiates et médiates de l'expression verbale.
- C - L'interaction professeur-étudiants	1. Dimension cognitive: interactions liées à la tâche d'enseignement (professeur) et/ou d'apprentissage (étudiants). 2. Dimension socio-émotive: interactions liées au climat socio-émotif dans et hors de la classe.

Suite à ce regroupement théorique des dimensions en six sections, un nouveau regroupement a été tenté en utilisant cette fois des données statistiques. Une technique statistique originale a été élaborée, technique présentant certaines ressemblances avec un "sociogramme", afin d'étudier les relations entre les 61 dimensions. Cette méthode a conduit à un nouveau regroupement qui a été décrit précédemment par le Dr Gagné; qu'il me suffise de vous rappeler l'appellation de ces sections.

TROISIÈME REGROUPEMENT

1. Encadrement
2. Versatilité
3. Rigoristisme
4. Altérocentrisme
5. Climat
6. Disponibilité

Ce dernier regroupement est maintenant utilisé pour l'interprétation des résultats; toutefois, d'autres recherches actuellement en cours visent à étudier, à l'aide d'autres techniques et d'autres approches théoriques, les relations entre les diverses dimensions.

En effet, plusieurs problèmes sont sous-jacents aux sections; mentionnons par exemple celui de la validité des mesures cumulées soit de section, soit globale.

Il faudra étudier, d'une part, jusqu'à quel point les notes d'insatisfaction des diverses dimensions ont une part de signification commune justifiant leur cumul en mesures synthétiques et, d'autre part, l'importance relative de chacune des dimensions du test ou en d'autres mots, la pondération des scores de dimensions lorsqu'on les cumule en scores synthétiques.

II - Structure de l'instrument

Examinons maintenant les problèmes liés à la structure de l'instrument. Le plan d'étude de ces problèmes comporte quatre étapes, à savoir

- l'étude du construit insatisfaction
- les problèmes de normalisation
- la purification des données
- la hiérarchisation des dimensions.

1ère étape: Etude du construit d'insatisfaction

Les postulats de base de ce construit vous ont été présentés tout à l'heure par le Dr Gagné. Ce concept est assez bien défini et s'appuie sur des théories bien établies; toutefois, nous examinons la possibilité de raffiner ce concept en introduisant la notion d'importance de l'insatisfaction pour chacun des étudiants qui répondent au test. Cette notion se greffe à celle de la hiérarchisation des dimensions; c'est pourquoi nous en discuterons plus loin.

3ème étape: Les problèmes de normalisation

Historiquement, le Service PERPE a tout d'abord fourni aux usagers du test des normes générales; puis il a ajouté à ces normes générales des normes de secteurs englobant plusieurs matières scolaires passablement hétérogènes. Lors de la dernière exploitation, ces normes de secteurs ont été remplacées par des normes de matières, ainsi, un professeur est comparé d'une part à l'ensemble des professeurs et d'autre part aux professeurs qui enseignent la même matière.

Toutefois, il nous apparaît essentiel de revenir à plusieurs niveaux la méthode de normalisation; en effet, plusieurs problèmes se posent, tant au niveau des groupes de référence auxquels le professeur est comparé qu'à celui des populations servant à construire les normes et qu'à celui des mesures qui sont normalisées.

2ème étape: La purification des données

Différents contrôles sont exercés sur les données brutes; en effet, les étudiants marginalement insatisfaits sont éliminés. Il en est de même pour ceux qui négligent de répondre à au moins ¾ des dimensions. De plus, une dimension est éliminée si un trop petit nombre d'étudiants y répondent. En plus de réexaminer les critères définis pour chacun de ces contrôles, nous porterons attention à la proportion de nos données qui en sont affectées.

4ème étape: La hiérarchisation des dimensions

Le test PERPE contient 61 dimensions qui ne sont pas toutes d'égale importance pour l'évaluation d'un cours. Plusieurs tentatives ont été faites afin de connaître l'importance relative accordée à chacune dimension par les professeurs et par les étudiants. Par exemple, lors d'une expérimentation, les étudiants et les professeurs devaient indiquer, sur une échelle en 5 points, l'importance qu'ils accordaient à chaque dimension; dans 83% des cas, les deux jugements n'étaient pas statistiquement différents. Quelques autres ten-

tatives artisanales ont fourni sensiblement les mêmes résultats, de sorte qu'on peut en inférer que l'entente entre les professeurs et les étudiants est assez forte. Toutefois, ces travaux n'ont pas encore fourni de réponse au problème de la pondération des dimensions.

C'est pourquoi nous avons l'intention, dans les mois qui vont suivre, d'approfondir ce sujet dans le double but de déterminer les dimensions les plus importantes de l'insatisfaction étudiante et de trouver une solution au problème de la pondération des résultats. Cette pondération est entrevue sous deux aspects, à savoir la pondération des dimensions en vue de constituer des scores globaux et la pondération de l'insatisfaction de chacun des étudiants sur chaque dimension de façon à tenir compte de l'importance que revêt, pour chaque étudiant, chacune des dimensions.

III - Qualités psychométriques de l'instrument

Examinons enfin les qualités psychométriques de l'instrument, à savoir la fidélité et la validité.

1ère étape: La fidélité du test

Tout instrument, s'il veut être d'une utilité quelconque, doit être fidèle. Les études faites jusqu'à maintenant ont montré que le test PERPE a un niveau de fidélité très satisfaisant; en effet, quelques études de type "split-half" ainsi que la comparaison de plusieurs profils d'un même professeur évalué par plusieurs classes tendent à démontrer la haute fidélité du test PERPE dans l'évaluation de la situation d'enseignement.

En effet, lorsqu'on scinde un groupe de répondants en deux, on observe une corrélation de l'ordre de 0.85 entre les deux groupes quant à l'évaluation du même professeur tandis que, lorsqu'on compare deux profils d'un même professeur évalué par deux classes différentes, on observe que ces deux profils se

ressemblent plus, dans le cas sur 10, que n'importe quel autre couple de profils de professeurs différents.

Toutefois, ces analyses n'ont pas été effectuées sur une grande échelle; il importera donc de voir jusqu'à quel point ces résultats peuvent se généraliser.

3ème étape: La validité du test

Une des qualités essentielles d'un instrument de mesure est d'être valide, c'est-à-dire de mesurer effectivement ce qu'il prétend mesurer. Nous ne pouvons toutefois qu'exposer très brièvement les principaux travaux effectués, ne pouvant condenser en quelques lignes un texte d'une centaine de pages[1].

Au plan de la validité de contenu, on a pu voir précédemment la rigueur scientifique sous-jacente à la construction des dimensions et à leur regroupement en sections. Qu'il suffise d'ajouter que chaque dimension à évaluer est opérationalisée en la définissant à l'aide d'exemples concrets; de plus, chaque dimension est unidimensionnelle en ce sens qu'elle ne mesure qu'une seule réalité. Au plan de la représentativité du contenu, la méthode de construction des dimensions, la revue de la littérature pertinente ainsi que les commentaires des usagers du test tendent à montrer que PERFE présente un éventail de dimensions qui couvrent très bien l'ensemble de la situation d'enseignement.

Au plan de la validité de construit, la seule variable théorique sous-jacente au test PERFE est le construit d'insatisfaction défini par l'écart entre la réalité et les attentes. Qu'il suffise d'ajouter à ce qui a déjà été mentionné à ce sujet que les dimensions semblent assez homogènes au plan de l'insatisfaction pour qu'il soit valide de parler d'insatisfaction globale.

Sur le plan de la validité de prédiction et de la validité concomitante qui, l'une et l'autre, exigent la présence d'un critère externe de

---

[1] CHABOT, Marie (1970). La validité du test PERFE, texte B-640, document 0570-02. Service PERFE

satisfaction ou d'insatisfaction, vous acquiescerez sûrement à l'affirmation qu'il n'existe aucun critère valide de ce type. Il est donc impossible actuellement d'obtenir des résultats vraiment significatifs sous ces rapports. Mentionnons toutefois que, lors d'une étude artisanale, on a pu montrer que l'insatisfaction globale d'un étudiant a une corrélation de 0.81 avec le rang qu'il attribue au cours évalué parmi tous ses autres cours. De telles données sont donc plutôt encourageantes.

Cependant, il sera de la plus haute importance de reprendre ces études à partir de l'ensemble des données.

En conclusion, qu'il me soit permis de vous rappeler, en plus de son importance immédiate, l'importance à long terme de ce programme d'analyse; en effet, dans la perspective de mise en marche d'un programme de construction de nouveaux instruments, ce dont le Dr Gagné vous entretiendra par la suite, ces analyses permettront d'identifier les lacunes de l'instrument actuel, d'où la possibilité de perfectionner nos techniques de mesure et d'offrir, dans l'avenir, des instruments toujours meilleurs.

. . . . . .